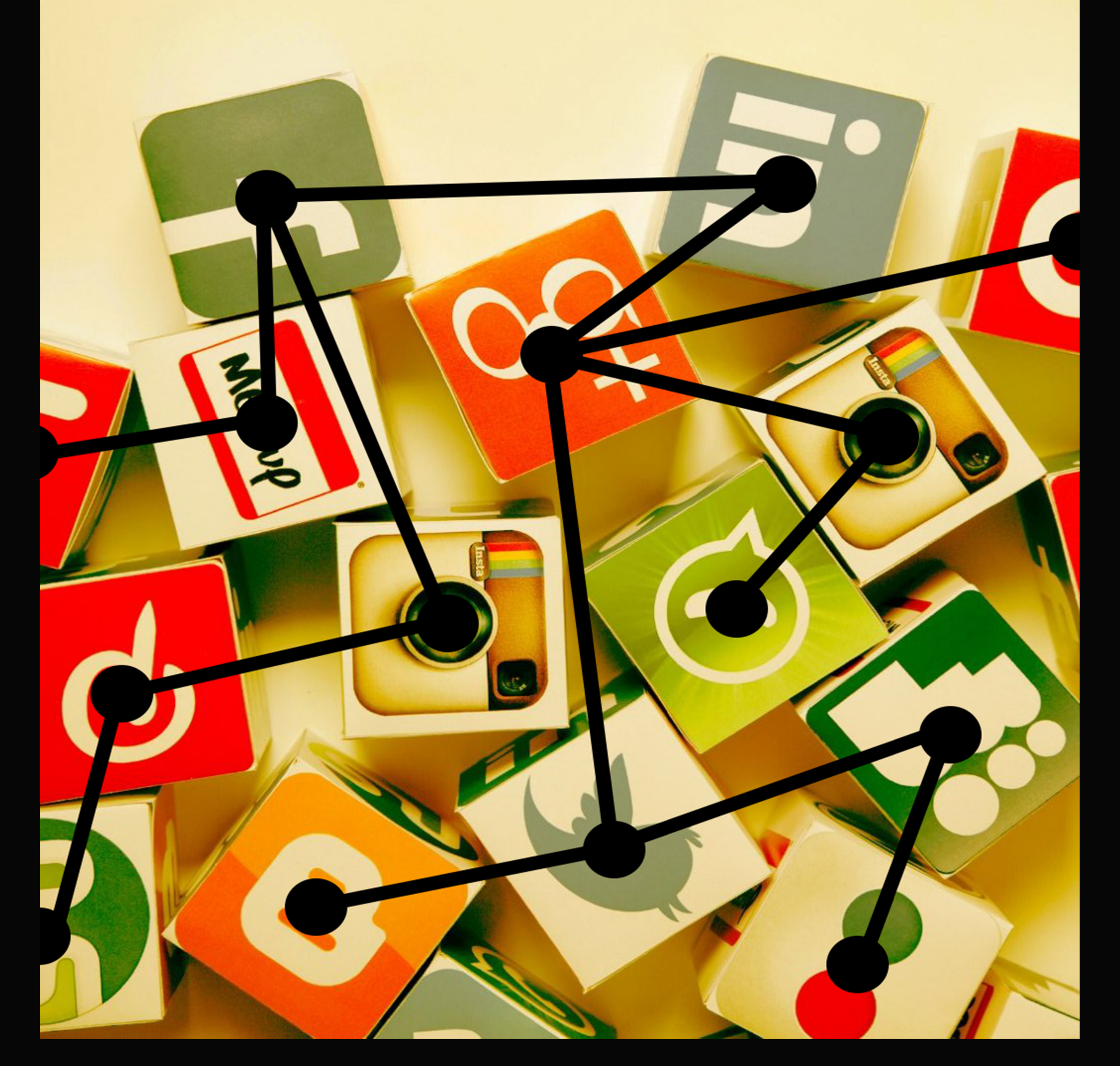
Análisis de redes sociales con R

Ingeniería Multimedia

Juan Manuel Ruiz Culiáñez



Introducción

El análisis de redes sociales es uno de los elementos más importantes en el entorno “social-media” a día de hoy. Podríamos definirlo como una aproximación práctica sobre el estudio de las relaciones entre las personas que las usan, pudiendo observar y analizar diferentes tendencias en función de diversos factores como círculos de amistad, nivel educativo, nacionalidad, etc.

Por lo tanto, a continuación vamos a realizar una serie de ejemplos sobre el tratado de diferentes informaciones que nos aportan las redes sociales, ya sean los intereses se cada usuario o el posible alcance de la información. Todo ello empleando como herramienta principal *RStudio*.

1- Análisis de gustos y publicidad personalizada

Las redes sociales más utilizadas (Facebook, Twitter, Instagram, etc) emplean algoritmos de recomendación muy avanzados para personalizar el contenido y la publicidad a cada usuario. La publicidad es uno de las principales fuentes de ingresos para este tipo de servicios, ya que pueden vender la información de los usuarios a grandes compañías para atraer de forma más eficaz a posibles clientes.

Uno de los métodos empleados para realizar esto es a través de las matrices de ponderación. Veamos un ejemplo:

Hay que definir un conjunto de personalidades que puedan ajustarse en la medida de lo posible a las preferencias de los usuarios. En este caso vamos a tratar con un joven interesado en las tecnologías de la comunicación, ocio digital y música.

Sabemos que existe ese perfil, pero, ¿Cómo lo distinguimos? Tanto los buscadores de las redes sociales como Google recopilan estos datos y los ponderan en función de la cantidad de veces que se busca y el tiempo de retención en cada página. Además sitios como Twitter también recopilan información de otros recursos como la localización o los *"Hashtag"*.

De esta forma, si nos ceñimos a los datos obtenidos de estas fuentes, podemos ponderar los diferentes temas según el interés que despertará al usuario:

(T="Tema")

- T1= Ropa (1.15)
- T2= Noticias del Corazón(1)
- T3= Tecnología y videojuegos (1.75)
- T4= Cine (1.5)
- T5= Música (1.5)
- T6= Automoción (1.25)

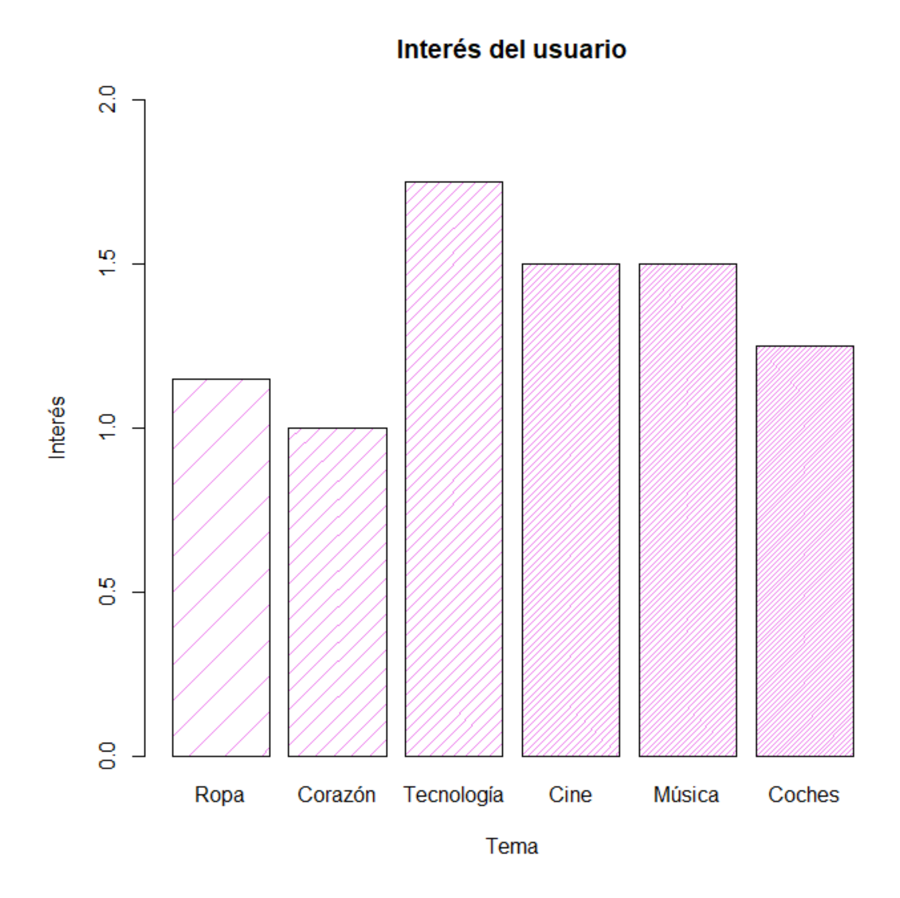
Crearíamos el Vector con estos datos en R:

```
V <- c(1.15,1,1.75,1.5,1.5,1.25)
```

Si visualizamos una gráfica:

```
barplot(V, col="violet", ylim=c(0,2), main="Interés del usuario",
xlab="Tema", ylab = "Interés", names.arg
=c("Ropa","Corazón","Tecnología","Cine","Música",
"Coches"), density=c(5,10,15,25,30,35))
```

El resultado sería el siguiente:



Ahora tenemos que filtrar toda la publicidad que tenemos disponible para que se muestre con prioridad la que más interese al usuario. Para ello tenemos que definir en forma de vector cada anuncio con la cantidad de interés que genera en cada tema:

(A= "Anuncio")

A1 (Anuncio Tele5)<- c(7,9,5,5,5,5)

A2 (Anuncio Videojuego) <- c(6,5,9,7,6,6)

A3 (Anuncio Película) <- c(6,6,7,9,8,6)

A4 (Anuncio Nuevo álbum) <- c(5,6,5,6,9,5)

A5 (Anuncio Tienda Ropa Grupos música) <- c(9,6,5,5,8,5)

A6 (Anuncio Coche Deportivo) <- c(6,6,7,5,5,9)

A continuación creamos una Matriz que agrupe todos los Vectores “A” que hemos definido:

```
MD <- rbind(A1, A2, A3, A4, A5, A6)
```

Y realizamos el producto:

```
RESULTANTE <- MD %*% V
```

Desde R podemos observar la siguiente Tabla:

	V1
A3	58.15
A2	54.65
A6	51.40
A5	50.85
A4	49.25
A1	47.05

De esta forma, hemos creado un sistema de filtrado con R para poder priorizar la publicidad en función del conjunto de intereses del usuario sobre cada tema.

Se incluye el archivo “**Publicidad.r**” para poder utilizarlo y editar libremente.

2- Visualización de relaciones entre los usuarios

Además de por tener la cantidad de usuarios activos más grande, Facebook es una red tan exitosa comercialmente por la cantidad de herramientas y facilidades que aporta a la comunidad para analizar audiencias y tendencias.

Es cierto que existen infinidad de herramientas ajenas a R para poder analizar diferentes parámetros, pero también encontramos algunas muy interesantes que crea la comunidad. Es el caso de RFacebook, librería de R creada por Pablo Barbera.

¿Qué nos permite esta herramienta? A grandes rasgos, RFacebook permite recopilar la información de Facebook en datos que pueda leer R. De esta forma, podemos tratar estos datos ya sea para averiguar tendencias con las propiedades de las matrices o visualizar con *plot* las interacciones entre la gente.

Lo que vamos a hacer exactamente con esta herramienta, es visualizar una gráfica (grafo) que muestre las relaciones de amistad que hay entre mis amigos y yo. Obviamente censuraré mi información personal y la de mis conocidos.

Una vez explicado esto, procedemos a comentar el procedimiento:

Preparar R

Debemos instalar las herramientas de desarrollo y sus librerías. Para ello introducimos:

```
install.packages("devtools")
library(devtools)
```

A continuación accedemos al *github* de Pablo Barbera para instalar todos los métodos de RFacebook:

```
install_github("Rfacebook", "pablobarbera",
subdir="Rfacebook")

require (Rfacebook)
```

Preparar Facebook (Developer Mode)

Para que sea posible la obtención de datos, debemos habilitar Facebook en modo desarrollador.

Accederemos a https://developers.facebook.com y activaremos el explorador de la API *Graph*, permitiendo el acceso a todos los datos posibles:

Seleccionar permisos v2.8 ×

Permisos de datos del usuario

email ×	user_hometown ×	user_religion_politics ×
publish_actions ×	user_likes ×	user_status ×
user_about_me ×	user_location ×	user_tagged_places ×
user_birthday ×	user_photos ×	user_videos ×
user_education_history ×	user_posts ×	user_website ×
user_friends ×	user_relationship_details ×	user_work_history ×
user_games_activity ×	user_relationships ×	

Eventos, grupos y páginas

ads_management ×	pages_messaging ×	read_page_mailboxes ×
ads_read ×	pages_messaging_payments ×	rsvp_event ×
business_management ×	pages_messaging_phone_number ×	user_events ×
manage_pages ×	pages_messaging_subscriptions ×	user_managed_groups ×
pages_manage_cta ×	pages_show_list ×	pages_manage_instant_articles
publish_pages ×		
user_actions.books ×	user_actions.music ×	user_actions.video ×
user_actions.fitness ×	user_actions.news ×	

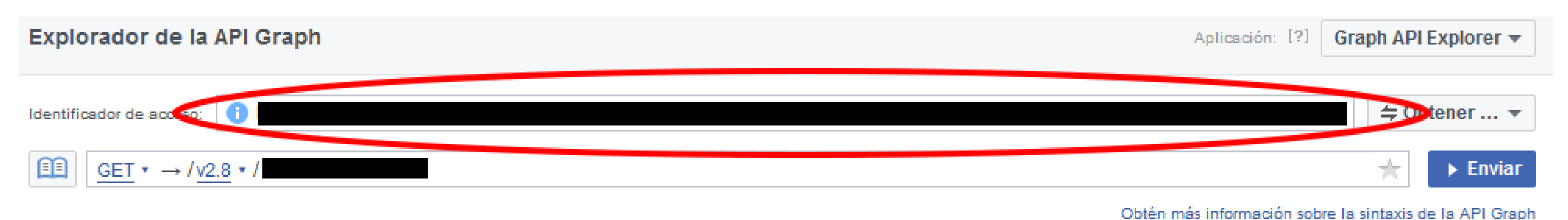
Otros

read_audience_network_insights ×	read_custom_friendlists ×	read_insights ×

El perfil público se incluye de forma predeterminada. Obtener identificador de acceso | Borrar | Cancelar

:

Una vez confirmado esto, podremos ver el “Token” que nos permitirá acceso a nuestro Facebook desde R. Se encuentra aquí:



Es importante deshabilitar todos estor permisos al terminar el proceso, así no ponemos en peligro nuestra cuenta con ellos.

Trabajar con R (Código muestreo grafo)

Una vez todo listo, podemos proceder a programar un breve Script en R para que muestre con un *plot* las relaciones entre nuestros conocidos.

Es importante tener instalado el paquete *“igraph”* en nuestro sistema.

El código es el siguiente:

```
tokenmio <- "PONER EL TOKEN AQUI"
yo <- getUsers("yo", token=tokenmio)
mis_amigos <- getFriends(token=tokenmio, simplify=TRUE)
mis_amigos_info <- getUsers(mis_amigos$id, token=tokenmio, private_info=TRUE)
mi_conexion <- getNetwork(tokenmio, format="adj.matrix")
solitarios <- rowSums(mi_conexion)==0 #Amigos que solo son amigos mios

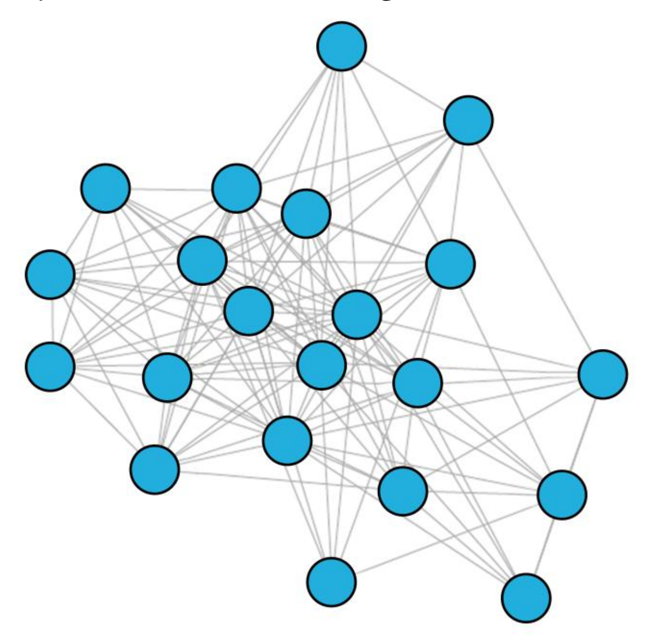
require(igraph)
mi_grafo <- graph.adjacency(mi_conexion[!solitarios,!solitarios])
layout <- layout.drl(mi_grafo,options=list(simmer.attraction=0))
plot(mi_grafo, vertex.size=2,
     #vertex.label=NA,
     vertex.label.cex=0.5,
     edge.arrow.size=0, edge.curved=TRUE,layout=layout)
```

Como podemos ver el proceso es muy sencillo. Simplemente nos hemos limitado a obtener la información de Facebook del perfil y de los amigos para posteriormente adaptarla a un plot.

También podemos comprobar los datos manualmente desde el apartado "Global Enviroment":

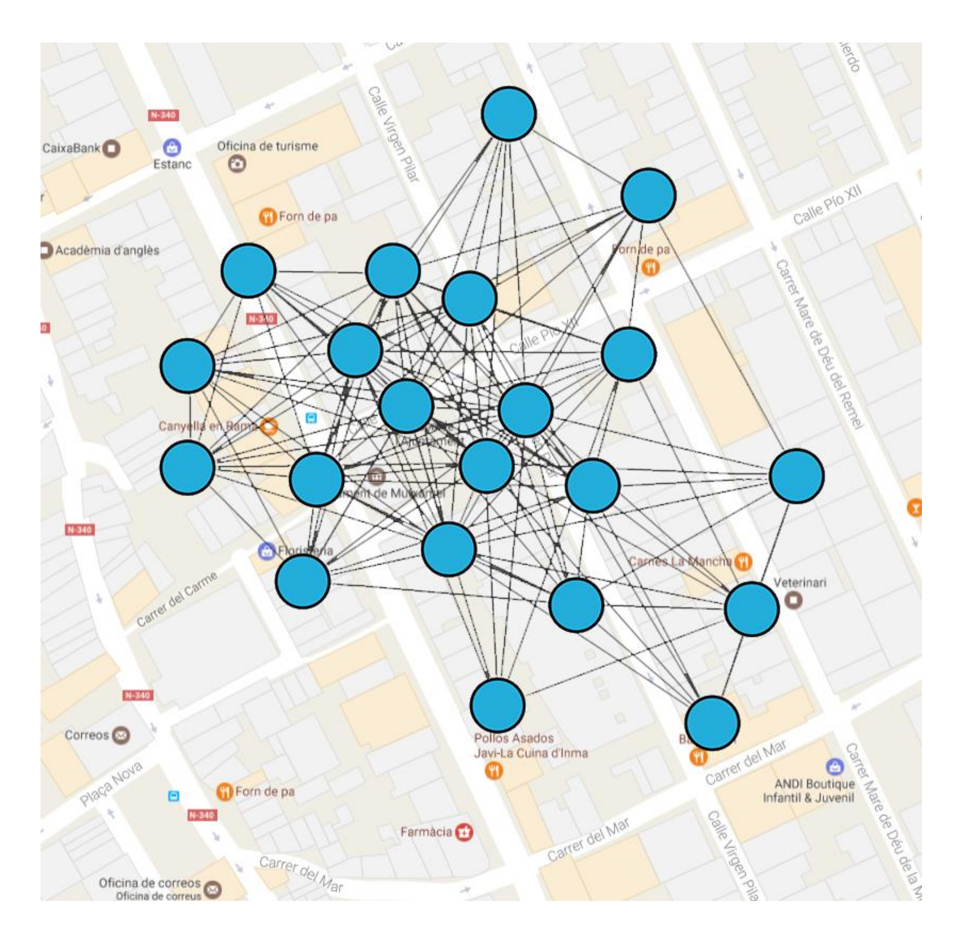
Data	
me	1 obs. of 10 variables
id : chr "[redacted]"	
name : chr "Juan Manuel Ruiz Culiañez"	
username : logi NA	
first_name : chr "Juan Manuel"	
middle_name: logi NA	
last_name : chr "Ruiz Culiañez"	
gender : chr "male"	
locale : chr "es_ES"	
likes : logi NA	
picture : logi NA	

En mi caso, no utilizo Facebook nunca y tengo muy pocos amigos agregados. Realizando el proceso obtendríamos un grafo similar al siguiente (Adyacencia, véase el código):



En el grafo no aparecen nombres ni datos de los usuarios por motivos de privacidad.

La información que acabamos de emplear es muy útil, pero todavía se puede refinar más. Por ejemplo, podemos ver cómo se relacionan estas personas si las situáramos en un mapa. Por ejemplo, si utilizamos Google Maps (Ciudad de Mutxamel, Alicante):



Observamos que las personas se relacionan más cuando más cerca están.

Esto se debe al principio de *Asortatividad* y *Disortatividad*:

- **Red Asortativa**: Nodos que están muy conectados tienden a estar rodeados de nodos con alta conectividad.

$$P(k'|k) = k'P(k')/\langle k \rangle$$

- **Red Disortativa**: Nodos que están poco conectados tienden a estar rodeados de nodos con baja conectividad.

Con esto, justificamos el hecho de que hay mayor probabilidad de que dos vecinos de un mismo nodo sean vecinos entre ellos (**Transitividad**). De esta forma, podemos ver que el grafo muestra la sociedad tal cual se comporta en la realidad (Círculos de conocidos).

Todo el código mostrado está disponible para su ejecución y modificación en el archivo "**RFacebook.r**".

Es necesario destacar que también ha aparecido una utilidad muy parecida para Twitter llamada "TwitteR". Su utilización es muy similar a la de RFacebook, pero he preferido trabajar con Facebook porque es una red social que utilizo menos y no me importa tanto exponer con este tipo de análisis.

3- Priorización de contenido

La gente visita las redes sociales para conocer más sobre la gente de su alrededor y para obtener información sobre noticias, deportes, humor, etc.

La mayoría de gente hace esto, pero depende de de cada uno la prioridad que da a cada apartado. Por ello, vamos a emplear la teoría de álgebra lineal con R para resolver este problema.

Las matrices de Transición (Cadenas de *Markov*) son muy útiles para averiguar la tendencia a un estado después de diferentes iteraciones. De esa forma, podemos aprovechar esta teoría para determinar los intereses de los usuarios según las veces que entre a cierta red social en un día.

Pongamos un ejemplo.

La matriz de transición es la siguiente:

Con ella, determinamos la probabilidad de que tras venir de un sitio, vayamos a otro.

XXXXXXX	Social	Noticias	Humor	Off-Topic
Social	0,40	0,30	0,30	0,20
Noticias	0,20	0,30	0,10	0,10
Humor	0,30	0,10	0,40	0,40
Off-Topic	0,10	0,30	0,20	0,30

En R lo definiríamos como:

```
A <- matrix(c(0.4, 0.2, 0.3, 0.1, 0.3, 0.3, 0.1, 0.3, 0.3,
0.1, 0.4, 0.2, 0.2, 0.1, 0.4, 0.3), nrow=4);
```

Después de esto, tenemos que indicar en una matriz estocástica la cantidad de gente que acude a cada apartado:

Social	0,40
Noticias	0,20
Humor	0,30
Off-Topic	0,10

En R lo definiríamos como:

```
B <- matrix(c(0.4, 0.2, 0.3, 0.1), nrow=4);
```

Para saber el resultado de la siguiente iteración, tan solo debemos de realizar la multiplicación de matriz fila-columna:

```
R2 <- A%*%B;
```

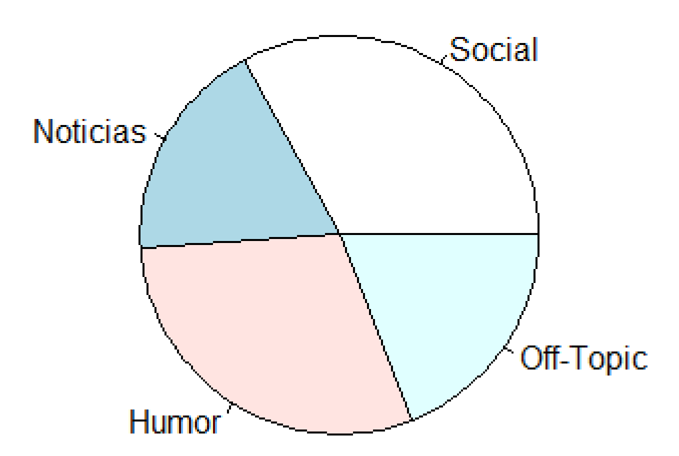
Obteniendo el siguiente resultado:

```
num [1:4, 1] 0.33 0.18 0.3 0.19
```

Si queremos ver los datos en forma de tarta gráfica, emplearemos el siguiente comando:

pie(R2, main="Interés del usuario",labels=c("Social","Noticias", "Humor", "Off-Topic"))

Obteniendo la siguiente gráfica:



Podemos realizar este proceso tantas veces queramos para poder averiguar cuáles son las tendencias del usuario según las veces que visita la red. (**R3 <- A%*%R2**)

Todo el código mostrado está disponible para su ejecución y modificación en el archivo "**Priorizacion.r**".

4- Información adicional

Si tratamos el análisis de redes sociales debemos tener en cuenta algunos elementos más allá del comportamiento de sus usuarios. No todas las redes ofrecen los mismos servicios y esto repercute en la cantidad de usuarios (popularidad) de las mismas.

Por ejemplo, si queremos socializar con la gente de nuestro alrededor, la más indicada es Facebook. Si queremos socializar a menor nivel pero también ver interacciones de los famosos que "admiramos" o ver cosas sobre cualquier tema, la más recomendada es Facebook. Instagram o Pinterest son las más eficaces para la imagen. Por último, aunque la gente no lo reconozca, aplicaciones de mensajería instantánea como Whatsapp o Telegram también son redes sociales por mucho que lo nieguen sus usuarios. Por ello debemos de tener mucho cuidado con lo que escribimos y exponemos en ellas.

Si prestamos atención al estudio anual del uso de las redes en 2016 del IAB (Asociación de la publicidad, el marketing y la comunicación digital en España), obtenemos que los datos sobre usos habituales son los siguientes:

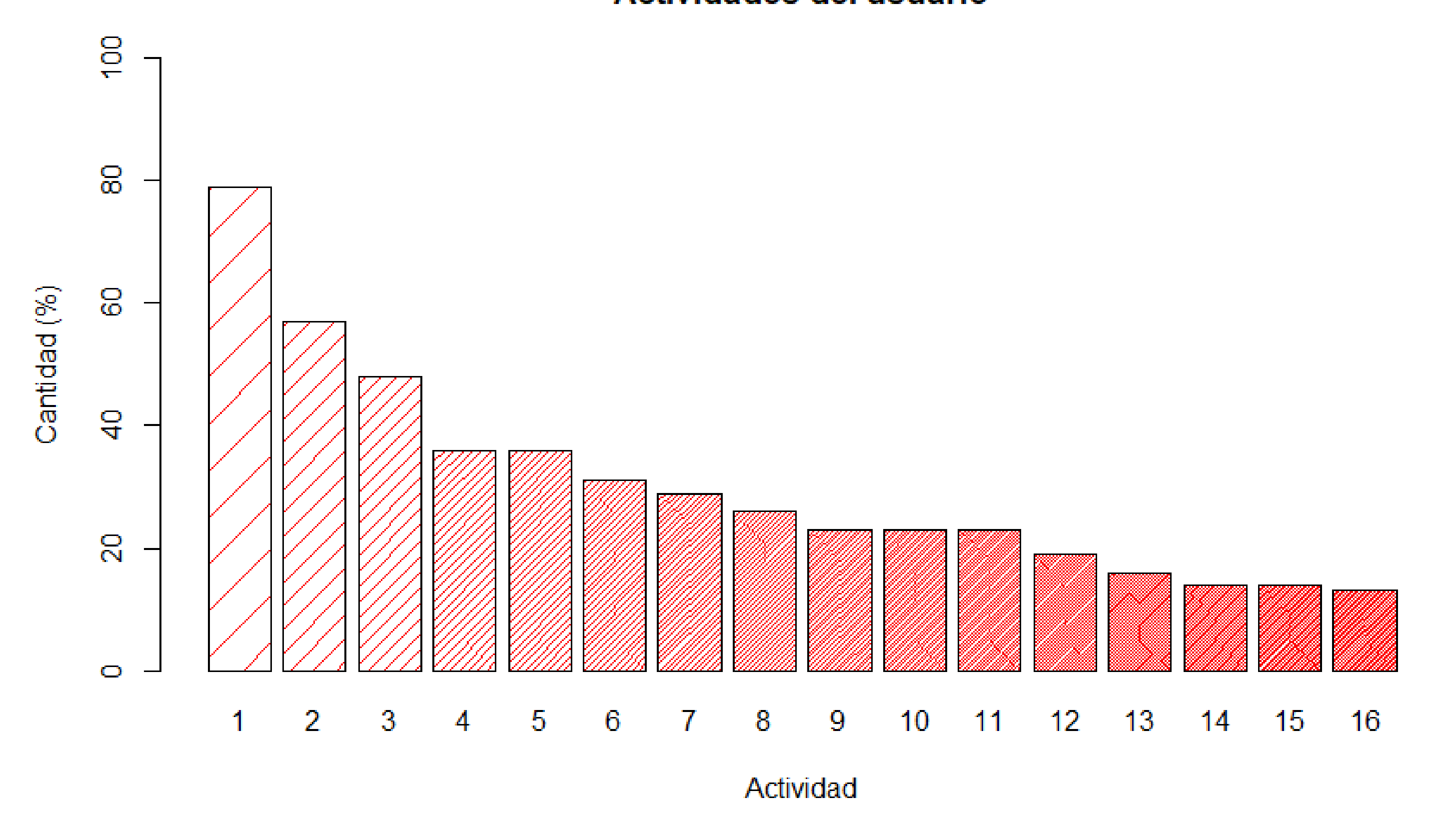
1- Chatear/Enviar Mensajes (79%)	9- Seguir a una marca (23%)
2- Ver videos, música (57%)	10- Hacer nuevos amigos (23%)
3- Ver qué hacen mis contactos (48%)	11- Participar en concursos (23%)
4- Publicar/Colgar Contenidos (36%)	12- Hablar de productos (19%)
5- Seguir a *influencers* (36%)	13- Comprar/Vender a contactos (16%)
6- Fines profesionales/estudios (31%)	14- Interactuar con la ubicación (14%)
7- Comentar la actualidad (29%)	15- Contactar al SAC (14%)
8- Jugar online (26%)	16- Crear eventos (13%)

Podemos emplear estos datos para mostrarlo de forma mucho más visual en R.

Representación con R:

```
V <- c(79, 57, 48, 36, 36, 31, 29, 26, 23, 23, 23,
        19, 16, 14, 14, 13);
barplot(V, col="red", ylim=c(0,100), main="Actividades del usuario", xlab="Actividad", ylab = "Cantidad (%)", names.arg =c("1","2","3","4","5", "6", "7","8","9","10","11", "12", "13","14","15","16"), density=c(5,10,15,25,30,35,40,45, 50,55,60,65, 70,75,80,85))
```

Obteniendo la siguiente gráfica:

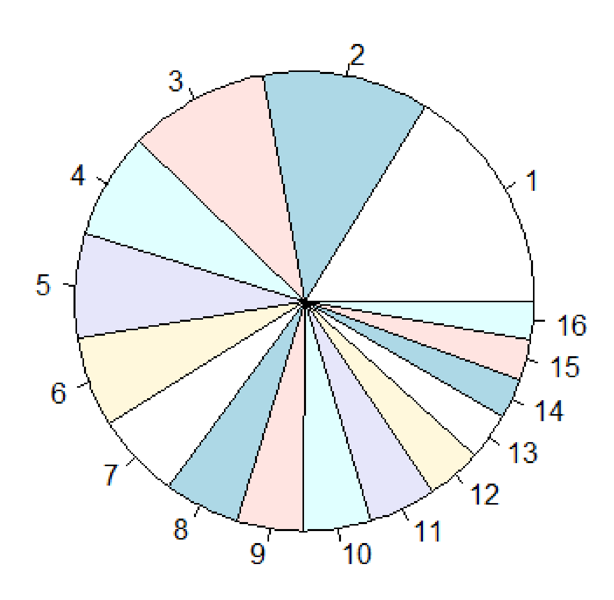


También podemos mostrarlo en forma de tarta:

Representación con R:

pie(V, main="Actividades del usuario", names.arg =c("1","2","3","4","5", "6", "7","8","9","10","11", "12", "13","14","15","16"));

Obteniendo la siguiente gráfica:



Véase el archivo "**InfoRedesPT1.r**" para obtener el Script en R.

Por otro lado, si analizamos la valoración y opinión de los usuarios sobre las redes sociales, obtendríamos los siguientes datos.

Whatsapp (8,5)	Facebook (7,5)	Google+ (6,1)
YouTube (8,1)	Pinterest (7,2)	XXXXXXXXXXXXXX
Spotify (7,7)	Twitter (7)	XXXXXXXXXXXXXX
Instagram (7,6)	LinkedIn (6,6)	XXXXXXXXXXXXXX
Telegram (7,5)	Line (6,4)	XXXXXXXXXXXXXX

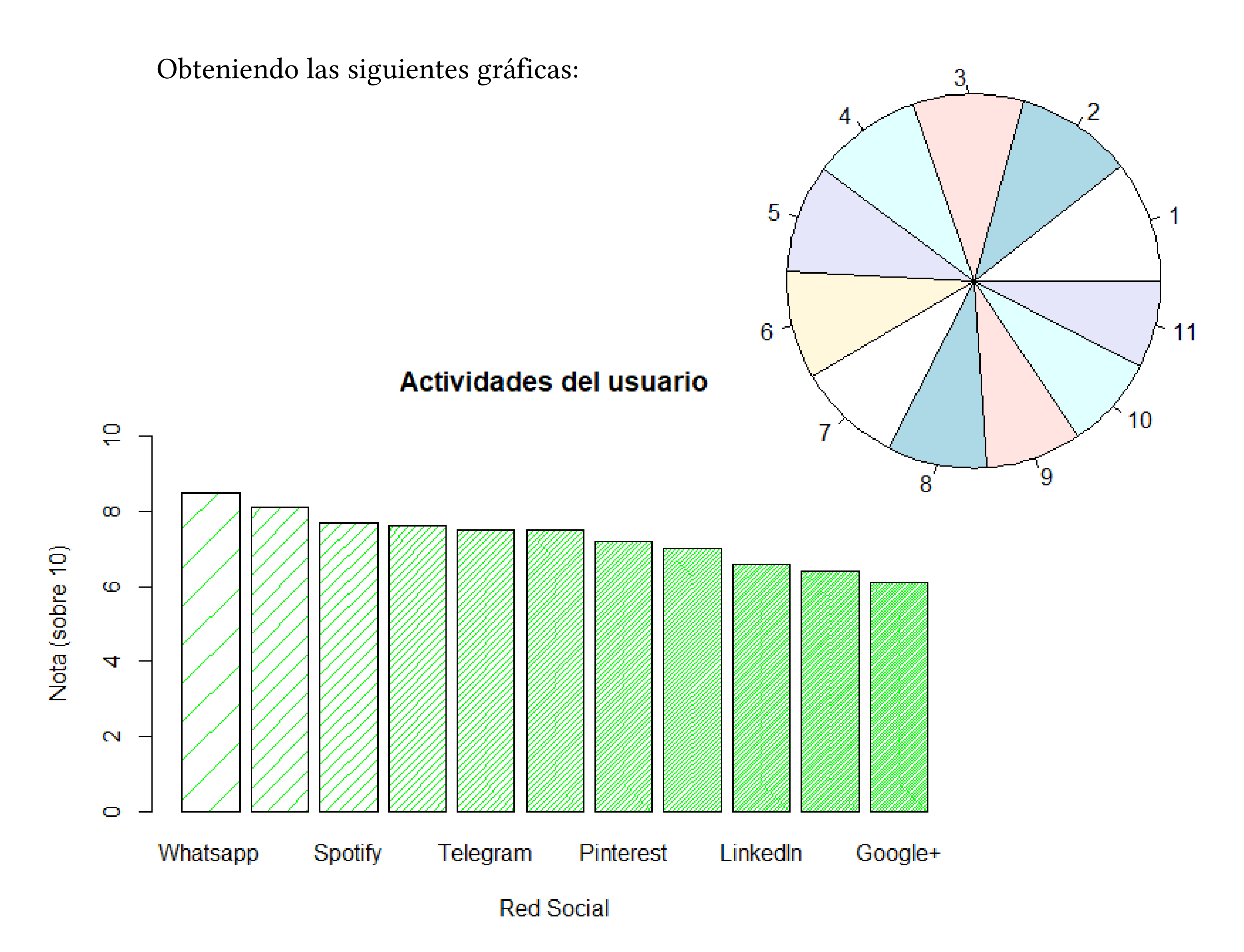
Podemos emplear estos datos para mostrarlo de forma mucho más visual en R.

Representación con R:

```
V <- c(8.5, 8.1, 7.7, 7.6, 7.5, 7.5, 7.2, 7, 6.6, 6.4, 6.1);

pie(V,  main="Notas de los usuarios", names.arg
=c("Whatsapp","YouTube","Spotify","Instagram",
"Telegram", "Facebook", "Pinterest","Twitter",
"LinkedIn","Line", "Google+"))
```

Obteniendo las siguientes gráficas:



Véase el archivo “**InfoRedesPT2.r**” para obtener el Script en R.